In the Claims:

1. (Original) A method of indexing a database of documents, comprising:

providing a vocabulary of n terms;

indexing the database in the form of a non-negative n x m index matrix V, wherein:

m is equal to the number of documents in the database;

n is equal to the number of terms used to represent the database; and

the value of each element v_{ij} of index matrix V is a function of the number of occurrences of the ith vocabulary term in the jth document;

factoring out non-negative matrix factors T and D such that $V \approx TD$; and

wherein T is an n x r term matrix, D is an r x m document matrix, and $r < nm/(n+m)$.

A2

2. (Original) The method of claim 1 further comprising deleting said index matrix V.

3. (Original) The method of claim 2 further comprising deleting said term matrix T.

4. (Original) The method of claim 1 wherein r is at least one order of magnitude smaller than n.

5. (Original) The method of claim 1 wherein r is from two to three orders of magnitude smaller than n.

6. (Original) The method of claim 1 wherein entries of said document matrix D falling below a predetermined threshold value t are set to zero.

7. (Original) The method of claim 2 wherein r is at least one order of magnitude smaller than n.

8. (Original) The method of claim 2 wherein r is from two to three orders of magnitude smaller than n.

9. (Original) The method of claim 2 wherein entries of said document matrix D falling below a predetermined threshold value t are set to zero.

10. (Original) The method of claim 3 wherein r is at least one order of magnitude smaller than n.

11. (Original) The method of claim 3 wherein r is from two to three orders of magnitude smaller than n.

12. (Original) The method of claim 3 wherein entries of said document matrix D falling below a predetermined threshold value t are set to zero.

13. (Currently Amended) A method of indexing a database of documents, comprising:

providing a vocabulary of n terms;

indexing the database in the form of a non-negative n x m index matrix V, wherein:

m is equal to the number of documents in the database;

n is equal to the number of terms used to represent the database; and

the value of each element v_{ij} of index matrix V is a function of the number of occurrences of the ith vocabulary term in the jth document;

A2

factoring out non-negative matrix factors T and D such that

$V \approx TD$. ~~The method of claim 1 wherein said factoring out of non-negative matrix factors T and D further comprises:~~

selecting a cost function and associated update rules from the group:

cost function $F = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[V_{ij} \log(TD)_{ij} - (TD)_{ij}\right]$ associated with

update rules $T_{ik} \leftarrow T_{ik} \sum_j \frac{V_{ij}}{(TD)_{ij}} D_{kj}$, $T_{ik} \leftarrow \frac{T_{ik}}{\sum_l T_{lk}}$, and $D_{kj} \leftarrow D_{kj} \sum_i T_{ij} \frac{V_{ij}}{(TD)_{ij}}$, cost function $F=\sum_{i=1}^{n}\sum_{j=1}^{m}\left[V_{ij}\log\frac{V_{ij}}{(TD)_{ij}}-(V_{ij})+(TD)_{ij}\right]$ associated with update rules $D_{kj}\leftarrow D_{kj}\frac{\sum_{i}\frac{T_{ik}V_{ij}}{(TD)_{ij}}}{\sum_{l}T_{lk}}$ and $T_{ik}\leftarrow T_{ik}\frac{\sum_{j}\frac{D_{kj}V_{ij}}{(TD)_{ij}}}{\sum_{h}D_{kh}}$, and

cost function $\|V-TD\|^{2}=\sum_{i=1}^{n}\sum_{j=1}^{m}\left(V_{ij}-(TD)_{ij}\right)^{2}$ associated with update rules $D_{kj}\leftarrow D_{kj}\frac{(T^{T}V)_{kj}}{(T^{T}TD)_{kj}}$ and $T_{ik}\leftarrow T_{ik}\frac{(VD^{T})_{ik}}{(TDD^{T})_{ik}}$; and

iteratively calculating said update rules so as to converge said cost function toward a limit until the distance between V and TD is reduced to or beyond a desired value.

A2

14. (Original) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for indexing a database of documents, said method steps comprising:

providing a vocabulary of *n* terms;

indexing the database in the form of a non-negative *n* x *m* index matrix *V*, wherein:

m is equal to the number of documents in the database;

n is equal to the number of terms used to represent the database; and

the value of each element v_{ij} of index matrix *V* is a function of the number of occurrences of the i^th vocabulary term in the j^th document;

factoring out non-negative matrix factors T and D such that

$$V \approx TD; \text{ and}$$

wherein T is an n x r term matrix, D is an r x m document matrix, and $r < nm/(n+m)$.

15. (Original) A database index, comprising:

an r x m document matrix D, such that

$$V \approx TD$$

wherein T is an n x r term matrix;

V is a non-negative n x m index matrix , wherein each of its m columns represents an j^th document having n entries containing the value of a function of the number of occurrences of a i^th term appearing in said j^th document; and

wherein T and D are non-negative matrix factors of V and $r < nm/(n+m)$; and

wherein each of the m columns of said document matrix D corresponds to said j^th document.

16. (Original) A method of information retrieval, comprising:

providing a query comprising a plurality of search terms;

providing a vocabulary of n terms;

performing a first pass retrieval through a first database representation and scoring m retrieved documents according to relevance to said query;

executing a second pass retrieval through a second database representation and scoring documents retrieved from said first pass retrieval so as to generate a final relevancy score for each document; and

wherein said second database representation comprises an r x m document matrix D, such that

$$V \approx TD$$

wherein T is an n x r term matrix;

V is a non-negative n x m index matrix , wherein each of its m columns represents an jth document having n entries containing the value of a function of the number of occurrences of a ith term of said vocabulary appearing in said jth document; and

wherein T and D are non-negative matrix factors of V and $r < nm/(n+m)$; and

wherein each of the m columns of said document matrix D corresponds to said jth document.

17. (Original) The method of claim 16 wherein said final relevancy score for any jth document is a function of said jth document s corresponding entry in said document matrix D and the corresponding entries in said document matrix D of the Γ top-scoring documents from said first pass retrieval.

18. (Original) The method of claim 17 wherein said relevancy score function for said jth document is proportional to a sum of cosine distances between said jth document s corresponding entry in said document matrix D and each of said

corresponding entries in said document matrix *D* of the Γ top- scoring documents from said first pass retrieval.

19. (Original) The method of claim 16 wherein *r* is at least one order of magnitude smaller than *n*.

20. (Original) The method of claim 16 wherein *r* is from two to three orders of magnitude smaller than *n*.

21. (Original) The method of claim 16 wherein entries of said document matrix *D* falling below a predetermined threshold value *t* are set to zero.

A2

22. (Original) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for information retrieval, said method steps comprising:

providing a query comprising a plurality of search terms;

providing a vocabulary of *n* terms;

performing a first pass retrieval through a first database representation and scoring *m* retrieved documents according to relevance to said query;

executing a second pass retrieval through a second database representation and scoring documents retrieved from said first pass retrieval so as to generate a final relevancy score for each document; and wherein said second database representation comprises an r x m document matrix D, such that

$$V \approx TD$$

wherein T is an n x r term matrix;

V is a non-negative n x m index matrix , wherein each of its m columns represents an jth document having n entries containing the value of a function of the number of occurrences of a ith term of said vocabulary appearing in said jth document; and

A2

wherein T and D are non-negative matrix factors of V and $r < nm/(n+m)$; and

wherein each of the m columns of said document matrix D corresponds to said jth document.